(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) Publication number: **0 538 877 A2**

(12) **EUROPEAN PATENT APPLICATION**

(21) Application number: **92118176.4**

(51) Int. Cl.[5]: **G10L 7/06**

(22) Date of filing: **23.10.92**

(30) Priority: **25.10.91 US 782669**

(43) Date of publication of application: **28.04.93 Bulletin 93/17**

(84) Designated Contracting States: **CH DE FR GB IT LI SE**

(71) Applicant: **MICOM COMMUNICATIONS CORP.**
4100 Los Angeles Avenue
Simi Valley, California 93063(US)

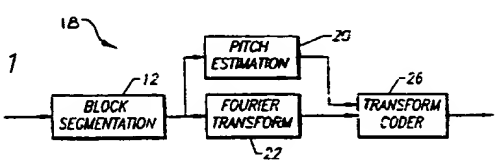
(72) Inventor: **Jain, Jaswant R.**
21638 Wohelo Trail
Chatsworth, California 91311(US)

(74) Representative: **Patentanwälte Grünecker, Kinkeldey, Stockmair & Partner**
Maximilianstrasse 58
W-8000 München 22 (DE)

(54) Voice coder/decoder and methods of coding/decoding.

(57) A new adaptive Fourier transform coder/decoder encodes periodic components of speech signals and decodes the encoded periodic components. The pitch frequency of voice signals in successive time frames at the voice coder may be determined as by (1) Cepstrum analysis (e.g. the time between successive peak amplitudes in each time frame), (2) harmonic gap analysis (e.g. the amplitude differences between the peaks and troughs of the peak amplitude signals of the frequency spectrum) (3) harmonic matching, (4) filtering of the frequency signals in successive pairs of time frames and the performance of (1), (2) and (3) on the filtered signals to provide pitch interpolation on the first frame in the pair and (5) pitch matching. The amplitude and phase of the pitch frequency and harmonic signals are determined by techniques refined relative to the prior art to provide amplitude and phase signals with enhanced resolution. Such amplitudes may be converted to a simplified digital form by (a) taking the logarithm of the frequency signals, (b) selecting the signal with the peak amplitude, (c) offsetting the amplitudes of the logarithmic signals relative to such peak amplitude, (d) companding the offset signals, (e) reducing the number of harmonics to a particular limit by eliminating alternate high frequency harmonics, (f) taking a discrete cosine transform of the remaining signals and (g) digitizing the transformed signals. If the pitch frequency has a continuity within particular limits in successive time frames, the phase difference of the signals between successive time frames is provided. At a displaced voice decoder, the signal amplitudes are determined by performing, in order, the inverse of steps (g) through (a). These signals and the signals representing pitch frequency and phase are processed to recover the voice signals.

FIG. 1



This invention relates to systems for, and methods of, encoding periodic components of voice signals in a voice coder for transmission to a voice decoder displaced from the voice coder. The invention also relates to a voice decoder for decoding the encoded voice signals transmitted from the voice encoder. The invention particularly relates to a voice encoder for encoding periodic components of voice signals with an enhanced resolution to provide for an optimal restoration of the voice signals at the voice decoder and also relates to a voice decoder for recovering the voice signals.

Microprocessors are used at a sending station to convert data to a digital form for transmission to a displaced position where the data in digital form is detected and converted to its original form. Although the microprocessors are small, they have enormous processing power. This has allowed sophisticated techniques to be employed by the microprocessor at the sending station to encode the data into digital form and to be employed by the microprocessor at the receiving station to decode the digital data and convert the digital data to its original form. The data transmitted may be through facsimile equipment at the transmitting and receiving stations and may be displayed as in a television set at the receiving station. As the processing power of the microprocessors has increased even as the size of the microprocessors has decreased, the sophistication in the encoding and decoding techniques, and the resultant resolution of the data at the receiving station, has become enhanced.

In recent years as the microprocessors have become progressively sophisticated in their ability to process data, it has become increasingly desirable to be able to transmit voice information in addition to data. For example, in telephone conferences, it has been desirable to transmit documents such as letters and written reports and analyses and to provide a discussion concerning such reports.

It has been found that it has been difficult to convert voice signals to a compressed digital form which can be transmitted to a receiving station to obtain a faithful reproduction of the speaker's voice at the receiving station. This results from the fact that the frequencies and amplitudes of a speaker's voice are constantly changing. This is even true during the time that the speaker is uttering a vowel, such as the letter "a", particularly since the duration of such vowels tends to be prolonged and speakers do not tend to talk in a monotone.

A considerable effort has been made, and a considerable amount of money has been expended, in recent years to provide systems for, and methods of, coding voice signals to a compressed digital form at a transmitting station, transmitting such digital signals to a receiving station and decoding such digital signals at the receiving station to reproduce the voice signals. As a result of such efforts and money expenditures, considerable progress has been made in providing a faithful reproduction of voice signals at the receiving station. However, in spite of such progress, a faithful reproduction of voice signals at the receiving station remains elusive. Listeners at the receiving station still do not hear the voice of a speaker at the transmitting station without inwardly feeling, or outwardly remarking, that there is a considerable distortion in the speaker's voice. This has tended to detract from the ability of the participants at the two (2) displaced stations to communicate meaningfully with each other.

This invention provides a system which converts voice signals into a compressed digital form in a voice coder to represent pitch frequency and pitch amplitude and the amplitudes and phases of the harmonic signals such that the voice signals can be reproduced at a voice decoder without distortion. The invention also provides a voice decoder which operates on the digital signals to provide such a faithful reproduction of the voice signals. The voice signals are coded at the voice coder in real time and are decoded at the voice decoder in real time.

In one embodiment of the invention, a new adaptive Fourier transform encoder encodes periodic components of speech signals and decodes the encoded signals. In the apparatus, the pitch frequency of voice signals in successive time frames at the voice coder may be determined as by (1) Cepstrum analysis (e.g. the time between successive peak amplitudes in each time frame, (2) harmonic gap analysis (e.g. the amplitude differences between the peaks and troughs of the peak amplitude signals of the frequency spectrum) (3) harmonic matching, (4) filtering of the frequency signals in successive pairs of time frames and the performance of steps (1), (2) and (3) on the filtered signals to provide pitch interpolation on the first frame in the pair, and (5) pitch matching.

The amplitude and phase of the pitch frequency signals and harmonic signals are determined by techniques refined relative to the prior art to provide amplitude and phase signals with enhanced resolution. Such amplitudes may be converted to a simplified digital form by (a) taking the logarithm of the frequency signals, (b) selecting the signal with the peak amplitude, (c) offsetting the amplitudes of the logarithmic signals relative to such peak amplitude, (d) companding the offset signals, (e) reducing the number of harmonics to a particular limit by eliminating alternate high frequency harmonics, (f) taking a discrete cosine transform of the remaining signals and (g) digitizing

the signals such transform. If the pitch frequency has a continuity within particular limits in successive time frames, the phase difference of the signals between successive time frames is provided.

At a displaced voice decoder, the signal amplitudes are determined by performing, in order, the inverse of steps (g) through (a). These signals and the signals representing pitch frequency and phase are processed to recover the voice signals without distortion.

In the drawings:

Figure 1 is a simplified block diagram of a system at a voice encoder for encoding voice signals into a digital form for transmission to a voice decoder;
Figure 2 is a simplified block diagram of a system at a voice decoder for receiving the digital signals from the voice encoder and for decoding the digital signals to reproduce the voice signals;
Figure 3 is a block diagram in increased detail of a portion of the voice encoder shown in Figure 1 and shows how the voice encoder determines and encodes the amplitudes and phases of the harmonics in successive time frames;
Figure 4 is a block diagram of another portion of the voice decoder and shows how the voice encoder determines the pitch frequency of the voice signals in the successive time frames;
Figure 5 is a block diagram of the voice decoder shown in Figure 2 and shows the decoding system in more detail than that shown in Figure 2;
Figure 6 is a schematic diagram of the voice signals to be encoded in successive time frames and further illustrates how the time frames overlap;
Figure 7 is a diagram schematically illustrating signals produced in a typical time frame to represent different frequencies after the voice signals in the time frame have been frequency transformed as by a Fourier frequency analysis;
Figure 8 illustrates the characteristics of a low pass filter for operating upon the frequency signals such as shown in Figure 7;
Figure 9 is a diagram schematically illustrating a spectrum of frequency signals after the frequency signals of Figure 7 have been passed through a low pass filter with the characteristics shown in Figure 8;
Figure 10 is a diagram illustrating one step involving the use of a Hamming window analysis in precisely determining the characteristics of each harmonic frequency in the voice signals in each time frame;
Figure 11 indicates the amplitude pattern of an individual frequency as a result of using the Hamming window analysis shown in Figure 10;
Figure 12 illustrates the techniques used to determine the amplitude and phase of each harmonic in the voice signals in each time frame with greater precision than in the prior art;
Figure 13 illustrates the relative amplitude values of the logarithms of the different harmonics in the voice signals in each time frame and the selection of the harmonic with the peak amplitude;
Figure 14 indicates the logarithmic harmonic signals of Figure 13 after the amplitudes of the different harmonics have been converted to indicate their amplitude difference relative to the peak amplitude shown in Figure 13;
Figure 15 schematically indicates the effect of a companding operation on the signals shown in Figure 14; and
Figure 16 illustrates how the frequency signals in different frequency slots or bins in each time frame are analyzed to provide voiced (binary "1") and unvoiced (binary "0") signals in such time frame.

In one embodiment of the invention, voice signals are indicated at 10 in Figure 6. As will be seen, the voice signals are generally variable with time and generally do not have a fully repetitive pattern. The system of this invention includes a block segmentation stage 12 (Figure 1) which separates the signals into time frames 14 (Figure 6) each preferably having a suitable time duration such as approximately thirty two milliseconds (32 ms.). Preferably the time frames 14 overlap by a suitable period of time such as approximately twelve milliseconds (12 ms.) as indicated at 16 in Figure 1. The overlap 16 is provided in the time frames 14 because portions of the voice signals at the beginning and end of each time frame 14 tend to become distorted during the processing of the signals in the time frame relative to the portions of the signals in the middle of the time frame.

The block segmentation stage 12 in Figure 1 is included in a voice coder generally indicated at 18 in Figure 1. A pitch estimation stage generally indicated at 20 estimates the pitch or fundamental frequency of the voice signals in each of the time frames 14 in a number of different ways each providing an added degree of precision and/or confidence to the estimation. The stages estimating the pitch frequency in different ways are shown in Figure 4.

The voice signals in each time frame 14 also pass to stage 22 which provides a frequency transform such as a Fourier frequency transform on the signals. The resultant frequency signals are generally indicated at 24 in Figure 7. The signals 24 in

each time frame 14 then pass to a coder stage 26. The coder stage 26 determines the amplitude and phase of the different frequency components in the voice signals in each time frame 14 and converts these determinations to a binary form for transmission to a voice decoder such as shown in Figures 2 and 5. The stages for providing the determination of amplitudes and phases and for converting these determinations to a form for transmission to the voice decoder of Figure 2 are shown in Figure 3.

Figure 4 illustrates in additional detail the pitch estimation stage 20 shown in Figure 1. The pitch estimation stage 20 includes a stage 30 for receiving the voice signals on a line 32 in a first one of the time frames 14 and for performing a frequency transform on such voice signals as by a Fourier frequency transform. Similarly, a stage 34 receives the voice signals on a line 36 in the next time frame 14 and performs a frequency transform such as by a Fourier frequency transform on such voice signals. In this way, the stage 30 performs frequency transforms on the voice signals in alternate ones of the successive time frames 14 and the stage 34 performs frequency transforms on the voice signals in the other ones of the time frames. The stages 30 and 34 perform frequency transforms such as Fourier frequency transforms to produce signals at different frequencies corresponding to the signals 24 in Figure 7.

The frequency signals from the stage 30 pass to a stage 38 which performs a logarithmic calculation on the magnitudes of these frequency signals. This causes the magnitudes of the peak amplitudes of the signals 24 to be closer to one another than if the logarithmic calculation were not provided. Harmonic gap measurements in a stage 40 are then provided on the logarithmic signals from the stage 38. The harmonic gap calculations involve a determination of the difference in amplitude between the peak of each frequency signal and the trough following the signal. This is illustrated in Figure 8 at 42 for a peak amplitude for one of the frequency signals 24 and at 44 for a trough following the peak amplitude 40. In determining the difference between the peak amplitudes such as the amplitude 42 and the troughs such as the trough 44, the positions in the frequency spectrum around the peak amplitude and the trough are also included in the determination. The frequency signal providing the largest difference between the peak amplitude and the following trough in the frequency signals 24 constitutes one estimation of the pitch frequency of the voice signals in the time frame 14. This estimation is where the peak amplitude of such frequency signal occurs.

As will be appreciated, female voices are higher in pitch frequency than male voices. This causes the number of harmonic frequencies in the voice signals of females to be lower than those in the voice signals of male voices. However, since the pitch frequency in the voice signals of a male is low, the spacing in time between successive signals at the pitch frequency in each time frame 14 may be quite long. Because of this, only two (2) or three (3) periods at the pitch frequency may occur in each time frame 14 for a male voice. This limits the ability to provide accurate determinations of pitch frequency for a male voice.

In providing a harmonic gap calculation, the stage 40 always provides a determination with respect to the voice frequencies of voices whether the voice is that of a male or a female. However, when the voice is that of a female, the stage 40 provides an additional calculation with particular attention to the pitch frequencies normally associated with female voices. This additional calculation is advantageous because there are an increased number of signals at the pitch frequency of female voices in each time frame 14, thereby providing for an enhancement in the estimation of the pitch frequency when an additional calculation is provided in the stage 40 for female voices.

The signals from the stage 40 for performing the harmonic gap calculation pass to a stage 46 for providing a pitch match with a restored harmonic synthesis. This restored harmonic synthesis will be described in detail subsequently in connection with the description of the transform coder stage 26 which is shown in block form in Figure 1 and in a detailed block form in Figure 3. The stage 46 operates to shift the determination of the pitch frequency from the stage 66 through a relatively small range above and below the determined pitch frequency to provide an optimal matching with such harmonic synthesis. In this way, the determination of the pitch frequency in each time frame is refined if there is still any ambiguity in this determination. As will be appreciated, a sequence of 512 successive frequencies can be represented in a binary sequence of nine (9) binary bits. Furthermore, the pitch frequency of male and female voices generally falls in this binary range of 512 discrete frequencies. As will be seen subsequently, the pitch frequency of the voice signals in each time frame 14 is indicated by nine (9) binary bits.

The signals from the stage 46 are introduced to a stage 48 for determining a harmonic difference. In the stage 48, the peak amplitudes of all of the odd harmonics are added to provide one cumulative value and the peak amplitudes of all of the even harmonics are added to provide another cumulative value. The two cumulative values are then compared. When the cumulative value for the even harmonics exceeds the cumulative value for the odd harmonics by a particular value such as approximately fifteen per cent (15%), the lowest

one of the even harmonics is selected as the pitch frequency. Otherwise, the lowest one of the odd harmonics is selected.

The voice signals on the lines 32 (for the alternate time frames 14) and 36 (for the remaining time frames 14) are introduced to a low pass filter 52. The filter 52 has characteristics for passing the full amplitudes of the signal components in the pairs of successive time frames with frequencies less than approximately one thousand hertz (1000Hz). This is illustrated at 54a in figure 8. As the frequency components increase above one thousand hertz (1000Hz), progressive portions of these frequency components are filtered. This is illustrated at 54b in Figure 8. As will be seen in Figure 8, the filter has a flat response 54a to approximately one thousand hertz (1000Hz) and the response then decreases relatively rapidly between a range of frequencies such as to approximately eighteen hundred hertz (1800Hz). The lowpass filtered signal is subsampled by a factor of two - i.e., alternate samples are discarded. This is consistent with the theory since the frequencies above 2000Hz have been nearly diminished.

The signals passing through the low pass filter 52 in Figure 4 are introduced to a stage 56 for providing a frequency transform such as a Fourier frequency transform. By filtering increasing amplitudes of the signals with progressive increases in frequency above one thousand Hertz (1000Hz), the frequency transformed signals generally indicated at 58 in Figure 9 are spread out more in the frequency spectrum than the signals in Figure 7. This may be seen by comparing the frequency spectrum of the signals produced in Figure 9 as a result of the filtering in comparison with the frequency spectrum in Figure 7. The spreading of the frequency spectrum in Figure 9 causes the resolution in the signals to be enhanced. For example, the frequency resolution may be increased by a factor of two (2).

The signals from the low pass filter 52 are also introduced to a stage 60 for providing a Cepstrum computation or analysis. Stages providing Cepstrum computations or analyses are well known in the art. In such a stage, the highest peak amplitude of the filtered signals in each pair of successive time frames 14 is determined. This signal may be indicated at 62 in Figure 6. The time between this signal 62 and a signal 64 with the next peak amplitude in the pair of successive time frames 14 may then be determined. This time is indicated at 66 in Figure 6. The time 66 is then translated into a pitch frequency for the signals in the pair of successive time frames 14.

The determination of the pitch frequency in the stage 60 is introduced to a stage 66 in Figure 4. The stage 66 receives the signals from a stage 68 which performs logarithmic calculations on the amplitudes of the frequency signals from the stage 56 in a manner similar to that described above for the stage 38. The stage 66 provides harmonic gap calculations of the pitch frequency in a manner similar to that described above for the stage 40. The stage 66 accordingly modifies (or provides a refinement in) the determination of the frequency from the stage 60 if there is any ambiguity in such determination. Alternatively, the stage 60 may be considered to modify (or provide a refinement in) the signals from the stage 66. As will be appreciated, there may be an ambiguity in the determination of the pitch frequency from the stage 60 if the time determination should be made from a different peak amplitude than the highest peak amplitude in the two (2) successive time frames or if the time between the successive peaks does not provide a precise indication of the pitch frequency.

As previously described, the stage 34 provides a frequency transform such as a Fourier frequency transform on the signals in the line 36 which receives the voice signals in the second of the two (2) successive time frames 14 in each pair. The frequency signals from the stage 34 pass to a stage 70 which provides a log magnitude computation or analysis corresponding to the log magnitude computations or analyses provided by the stages 38 and 68. The signals from the stage 70 in turn pass to the stage 66 to provide a further refinement in the determination of the pitch frequency for the voice signals in each pair of two (2) successive time frames 14.

The signals from the stage 66 pass to a stage 74 which provides a pitch match with a restored harmonic synthesis. This restored harmonic synthesis will be described in detail subsequently in connection with the description of the transform coder stage 26 which is shown in block form in Figure 1 and in a detailed block form in Figure 3. The pitch match performed by the stage 74 corresponds to the pitch match performed by the stage 46. The stage 74 operates to shift the determination of the pitch frequency from the stage 66 through a relatively small range above and below this determined pitch frequency to provide an optimal matching with such harmonic synthesis. In this way, the determination of the pitch frequency in each time frame is refined if there is still any ambiguity in this determination.

A stage 78 receives the refined determination of the pitch frequency from the stage 74. The stage 78 provides a further refinement in the determination of the pitch frequency in each time frame if there is still any ambiguity in such determination. The stage 78 operates to accumulate the sum of the amplitudes of all of the odd harmonics in the frequency transform signals obtained by the stage

74 and to accumulate the sum of the amplitudes of all of the even harmonics in such frequency transform. If the accumulated sum of all of the even harmonics exceeds the accumulated sum of all of the odd harmonics by a particular magnitude such as fifteen percent (15%) of the accumulated sum of the odd harmonics, the lowest frequency in the even harmonics is chosen as the pitch frequency. If the accumulated sum of the even harmonics does not exceed the accumulated sum of the odd harmonics by this threshold, the lowest frequency in the odd harmonics is selected as the pitch frequency. The operation of the harmonic difference stage 78 corresponds to the operation of the harmonic difference stage 48.

The signals from the stage 78 pass to a pitch interpolation stage 80. The pitch interpolation stage 80 also receives through a line 82 signals which represent the signals obtained from the stage 78 for one (1) previous frame. For example, if the signals passing to the stage 80 from the stage 78 represent the pitch frequency determined in time frames 1 and 2, the signals on the line 82 represent the pitch frequency determined for the frame 0. The stage 80 interpolates between the pitch frequency determined for the time frame 0 and the time frames 1 and 2 and produces information representing the pitch frequency for the time frame 1. This information is introduced to the stage 40 to refine the determination of the pitch frequency in that stage for the time frame 1.

The pitch interpolation stage 80 also employs heuristic techniques to refine the determination of pitch frequency for the time frame 1. For example, the stage 80 may determine the magnitude of the power in the frequency signals for low frequencies in the time frames 1 and 2 and the time frame 0. The stage 80 may also determine the ratio of the cumulative magnitude of the power in the frequency signals at low frequencies (or the cumulative magnitude of the amplitudes of such signals) in such time frames relative to the cumulative magnitude of the power (or the cumulative magnitude of the amplitudes) of the high frequency signals in such time frames. These factors, as well as other factors, may be used in the stage 80 in refining the pitch frequency for the time frame 1.

The output from the pitch interpolation stage 80 is introduced to the harmonic gap computation stage 40 to refine the determination of the pitch frequency in the stage 38. As previously described, this determination is further refined by the pitch match stage 46 and the harmonic difference stage 48. The output from the harmonic difference stage 48 indicates in nine (9) binary bits the refined determination of the pitch frequency for the time frame 1. These are the first binary bits that are transmitted to the voice decoder shown in Figure 2 to indicate to the voice decoder the parameters identifying the characteristics of the voice signals in the time frame 1. In like manner, the harmonic difference stage 78 indicates in nine (9) binary bits the refined estimate of the pitch frequency for the time frame 2. These are the first binary bits that are transmitted to the voice decoder shown in Figure 2 to indicate the parameters of the voice signals in the time frame 2. As will be appreciated, the system shown in Figure 4 and described above operates in a similar manner to determine and code the pitch frequency in successive pairs of time frames such as time frames 3 and 4, 5 and 6, etc.

The transform coder 26 in Figure 1 is shown in detail in Figure 3. The transform coder 26 includes a stage 86 for determining the amplitude and phase of the signals at the fundamental (or pitch) frequency and the amplitude and phase of each of the harmonic signals. This determination is provided in a range of frequencies to approximately four KiloHertz (4 KHz) bandwidth. The determination is limited to approximately four KiloHertz (4 KHz) because the limit of four thousand hertz (4Kz) corresponds to the limit of frequencies encountered in the telephone network as a result of adapted standards.

As a first step in determining the amplitude and phase of the pitch frequency and the harmonics in each time frame 14, the stage 86 divides the frequency range to four thousand Hertz (4000Hz) into a number of frequency blocks such as thirty (32). The stage 86 then divides each frequency block into a particular number of grids such as approximately sixteen (16). Several frequency blocks 96 and the grids 98 for one of the frequency blocks are shown in Figure 12. The stage 86 knows, from the determination of the pitch frequency in each time frame 14, the frequency block in which each harmonic frequency is located. The stage 86 then determines the particular one of the sixteen (16) grids in which each harmonic is located in its respective frequency block. By precisely determining the frequency of each harmonic signal, the amplitude and phase of each harmonic signal can be determined with some precision, as will be described in detail subsequently.

As a first step in determining with some precision the frequency of each harmonic signal in the Fourier frequency transform produced in each time frame 14, the stage 86 provides a Hamming window analysis of the voice signals in each time frame 14. A Hamming window analysis is well known in the art. In a Hamming window analysis, the voice signals 92 (Figure 10) in each time frame 14 are modified as by a curve having a dome-shaped pattern 94 in Figure 10. As will be seen, the dome-shaped pattern 94 has a higher am-

plitude with progressive positions toward the center of the time frame 14 then toward the edges of the time frame. This relative de-emphasis of the voice signals at the opposite edges of each time frame 14 is one reason why the time frames are overlapped as shown in Figure 6.

When the Hamming pattern 94 is used to modify the voice signals in each time frame 14 and a Fourier transform is made of the resultant pattern for an individual frequency, a frequency pattern such as shown in Figure 11 is produced. This frequency pattern may be produced for one of the sixteen (16) grids in the frequency block in which a harmonic is determined to exist. Similar frequency patterns are determined for the other fifteen (15) grids in the frequency block. The grid which is nearest to the location of a given harmonic is selected. By determining the particular one of the sixteen (16) grids in which the harmonic is located, the frequency of the harmonic is selected with greater precision than in the prior art.

In this way, the amplitude and phase are determined for each harmonic in each time frame 14. The phase of each harmonic is encoded for each time frame 14 by comparing the harmonic frequency in each time frame 14 with the harmonic frequency in the adjacent time frames. As will be be appreciated, changes in the phase of a harmonic signal result from changes in frequency of that harmonic signal. Since the period in each time frame 14 is relatively short and since there is a time overlap between adjacent time frames, any changes in pitch frequency in successive time frames may be considered to result in changes in phase.

As a result of the analysis as discussed above, pairs of signals are generated for each harmonic frequency, one of these signals representing amplitude and the other representing phase. These signals may be represented as $a_1\phi_1$, $a_2\phi_2$, $a_3\phi_3$, etc.

In this sequence

a_1, a_2, a_3, etc. represent the amplitudes of the signals at the fundamental frequency and the second, third, etc. harmonics of the pitch frequency signals in each time frame; and

ϕ_1, ϕ_2, ϕ_3, etc. represent the phases of the signals at the fundamental frequency and the second, third, etc. harmonics in each time frame 14.

Although the amplitude values a_1, a_2, a_3, etc., and the phase values ϕ_1, ϕ_2, ϕ_3, etc. may represent the parameters of the signals at the fundamental pitch frequency and the different harmonics in each time frame 14 with some precision, these values are not in a form which can be transmitted from the voice coder 18 shown in Figure 1 to a Voice decoder generally indicated at 100 in Figure 2. The circuitry shown in Figure 3 provides a conversion of the amplitude values a_1, a_2, a_3, etc., and the phase values ϕ_1, ϕ_2, ϕ_3, etc. to a meaningful binary form for transmission to the voice decoder 100 in Figure 2 and for decoding at the voice decoder.

To provide such a conversion, the signals from the harmonic analysis stage 86 in Figure 3 are introduced to a stage 104 designated as "spectrum shape calculation". The stage 104 also receives the signals from a stage 102 which is designated as "get band amplitude". The input to the stage 102 corresponds to the input to the stage 86. The stage 102 determines the frequency band in which the amplitude of the signals occurs.

As a first step in converting the amplitudes a_1, a_2, a_3, etc., to meaningful and simplified binary values for transmission to the voice decoder 100, the logarithms of the amplitude values a_1, a_2, a_3, etc., are determined in the stage 104 in Figure 3. Taking the logarithm of these amplitude values is desirable because the resultant values become compressed relative to one another without losing their significance with respect to one another. The logarithms can be with respect to any suitable base value such as a base value of two (2) or a base value of ten (10).

The logarithmic values of amplitude are then compared in the stage 104 in Figure 3 to select the peak value of all of these amplitudes. This is indicated schematically in Figure 13 where the different frequency signals and the amplitudes of these signals are indicated schematically and the peak amplitude of the signal with the largest amplitude is indicated at 106. The amplitudes of all of the other frequency signals are then scaled with the peak amplitude 106 as a base. In other words, the difference between the peak amplitude 106 and the magnitude of each of the remaining amplitude values a_1, a_2, a_3, etc., is determined. These difference values are indicated schematically at 108 in Figure 14.

The difference values 108 in Figure 14 are next companded. A companding operation is well known in the art. In a companding operation, the difference values shown in Figure 14 are progressively compressed for values at the high end of the amplitude range. This is indicated schematically at 110 in Figure 15. In effect, the amplitude values closest to the peak values in Figure 13 are emphasized by the companding operation relative to the amplitudes of low value in Figure 13.

As the next step in converting the amplitude values a_1, a_2, a_3, etc., to a meaningful and simplified binary form, the number of such values is limited in the stage 104 to a particular value such as forty five (45) if the amplitude values exceed forty five (45). This limit is imposed by disregarding the harmonics having the highest frequency

values. Disregarding the harmonics of the highest frequency does not result in any deterioration in the faithful reproduction of sound since most of the information relating to the sound is contained in the low frequencies.

As a next step, the number of harmonics is limited in the stage 104 to a suitable number such as sixteen (16) if the number of harmonics is between sixteen (16) and twenty (20). This is accomplished by eliminating alternate ones of the harmonics at the high end of the frequency range if the number of harmonics is between sixteen (16) and twenty (20). If the number of harmonics is less than sixteen (16), the harmonics are expanded to sixteen (16) by pairing successive harmonics at the upper frequency end to form additional harmonics between the paired harmonics and by interpolating the amplitudes of the additional harmonics in accordance with the amplitudes of the paired harmonics.

In like manner, if the number of harmonics is greater than twenty four (24), alternate ones of the harmonics are eliminated at the high end of the frequency range until the number of harmonics is reduced to twenty four (24). If the number of harmonics is between twenty one (21) and twenty four (24), the number of harmonics is increased to twenty four (24) by pairing successive harmonics at the upper frequency end to form additional harmonics between the paired harmonics and by interpolating the amplitudes of the additional harmonics in accordance with the amplitudes of the paired harmonics.

After the number of harmonics has been limited to sixteen (16) or twenty four (24) depending upon the number of harmonics produced in the Fourier frequency transform, a discrete cosine transform is provided in the stage 104 on the limited number of harmonics. The discrete cosine transform is well known to be advantageous for compression of correlated signals such as in a spectrum shape. The discrete cosine transform is taken over the full range of sixteen (16) or twenty four (24) harmonics. This is different from the prior art because the prior art obtains several discrete cosine transforms of the harmonics, each limited to approximately eight (8) harmonics. However, the prior art does not limit the total number of frequencies in the transform such as is provided in the system of this invention when the number is limited to sixteen (16) or twenty four (24).

The results obtained from the discrete cosine transform discussed in the previous paragraph are subsequently converted by a stage 110 to a particular number of binary bits to represent such results. For example, the results may be converted to forty eight (48), sixty four (64) or eighty (80) binary bits. The number of binary bits is preselected so that the voice decoder 100 will know how to decode such binary bits. In coding the results of the discrete cosine transform, a greater emphasis is preferably placed on the low frequency components of the discrete cosine transform relative to the high frequency components. For example, the number of binary bits used to indicate the successive values from the discrete cosine transform may illustratively be a sequence 5, 5, 4, 4, 3, 3, 3...2, 2..., , 0, 0, 0. In this sequence, each successive number from the left represents a component of progressively increasing frequency. The 48, 64 or 80 binary bits representing the results of the discrete cosine transform are transmitted to the voice decoder 100 in Figure 2 after the transmission of the nine (9) binary bits representing the pitch or fundamental frequency.

A stage 112 in Figure 3 receives the signals representing the discrete cosine transform from the stage 104 and reconstructs these signals to a form corresponding to the Fourier frequency transform signals introduced to the stage 86. As a first step in this reconstruction, the stage 112 receives the signals from the stage 104 and provides an inverse of a discrete cosine transform. The stage 112 then expands the number of harmonics to coincide with the number of harmonics in the Fourier frequency transform signals introduced to the stage 86. The stage 112 does this by interpolating between the amplitudes of successive pairs of harmonics in the upper end of the frequency range. The stage 112 then performs a decompanding operation which is the inverse of the companding operation performed by the stage 110. The signals are now in a form corresponding to that shown in Figure 14.

To convert the signals to the form shown in Figure 13, a difference is determined between the peak amplitude 106 shown in Figure 13 for each harmonic and the amplitude shown in Figure 14 for such harmonic. The resultant amplitudes correspond to those shown in Figure 13, assuming that each step in the reconversion provided in the stage 112 provides ideal calculations. The signals corresponding to those shown in Figure 13 are then processed in the stage 112 to remove the logarithmic values and to obtain Fourier frequency transform signals corresponding to those introduced to the stage 86.

The reconstructed Fourier frequency transform signals from the stage 112 are introduced to a stage 116. The Fourier frequency transform signals passing to the stage 86 are also introduced to the stage 116 for comparison with the reconstructed Fourier frequency transform signals in the stage 112. To provide this comparison, the Fourier frequency transform signals from each of the stages 86 and 112 are considered to be disposed in twelve (12) frequency slots or bins 118 as shown in

Figure 16. Each of the frequency slots or bins 118 has a different range of frequencies than the other frequency slots or bins. The number of frequency slots or bins is arbitrary but twelve (12) may be preferable. It will be appreciated that more than one (1) harmonic may be located in each time slot or bin 118.

The stage 116 compares the amplitudes of the Fourier frequency transform signals from the stage 112 in each frequency slot or bin 118 and the signals introduced to the stage 86 for that frequency slot or bin. If the amplitude match is within a particular factor for an individual one of the time slot or bin 118, the stage 116 produces a binary "1" for that time slot or bin. If the amplitude match is not within the particular factor for an individual time slot or bin 118, the stage 116 produces a binary "0" for that time slot or bin. The particular factor may depend upon the pitch frequency and upon other quality factors.

Figure 16 illustrates when a binary "1" is produced in a time slot or bin 118 and when a binary "0" is produced in a time slot or bin 118. As will be seen, when the correlation between the signals in the stages 86 and 112 is high as indicated by a signal of large amplitude, a binary "1" is produced in a time slot or bin 118. However, when the correlation is low as indicated by a signal of low amplitude, a binary "0" is produced for a time slot or bin 118. In effect, the stage 116 provides a binary "1" only in the frequency slots or bins 118 where the stage 104 has been successful in converting the frequency indications in the stage 86 to a form closely representing the indications in the stage 86. In the time slots or bins 118 where such conversion has not been successful, the stage 116 provides a binary "0".

Some post processing may be provided in the stage 116 to reconsider whether the binary value for a time slot or bin 118 is a binary "1" or a binary "0". For example, if the binary values for successive time slots or bins is "000100", the binary value of "1" in this sequence in the time frame 114 under consideration may be reconsidered in the stage 116 on the basis of heuristics. Under such circumstances, the binary value for this time slot or bin in the adjacent time frames 14 could also be analyzed to reconsider whether the binary value for this time slot or bin in the time frame 14 under consideration should actually be a binary "0" rather than a binary "1". Similar heuristic techniques may also be employed in the stage 116 to reconsider whether the binary value of "0" in the sequence of 11101 should be a binary "1" rather than a binary "0".

The twelve (12) binary bits representing a binary "1" or a binary "0" in each of the twelve (12) time slots or bins (118) in each time frame 14 are introduced to the stage 110 in Figure 3 for transmission to the voice decoder 100 shown in Figure 1. These twelve (12) binary bits in each time frame may be produced immediately after the nine (9) binary bits representing the pitch frequency and may be followed by the 48, 64 or 80 binary bits representing the amplitudes of the different harmonics. A binary "1" in any of these twelve (12) time bins or slots 118 may be considered to represent voiced signals for such time bin or slot. A binary "0" in any of these twelve (12) time bins or slots 118 may be considered to represent unvoiced signals for such time bin or slot. For a time bin or slot where unvoiced signals are produced, the amplitude of the harmonic or harmonics in such time bin or slot may be considered to represent noise at an average of the amplitude levels of the harmonic or harmonics in such time slot or bin.

The binary value representing the voiced (binary "1") or unvoiced (binary "0") signals from the stage 116 are introduced to the stage 104. For the time slots or bins 118 where a binary "1" has been produced by the stage 116, the stage 104 produces binary signals representing the amplitudes of the signals in the time slots or bins. These signals are encoded by the stage 110 and are transmitted through a line 124 to the voice decoder shown in Figure 2. When a binary "0" is produced by the stage 116 for a time slot or bin 118, the stage 104 produces "noise" signals having an amplitude representing the average amplitude of the signals in the time slot or bin. These signals are encoded by the stage 110 into binary form and are transmitted through the line 124 to the voice decoder.

The phase signals ϕ_1, ϕ_2, ϕ_3, etc. for the successive harmonics in each time frame 14 are converted in a stage 120 in Figure 3 to a form for transmission to the voice decoder 100. If the phase of the signals for a harmonic has at least a particular continuity in a particular time frame 14 with the phase of the signals for the harmonic in the previous time frame, the phase of the signal for the harmonic in the particular time frame is predicted from the phase of the signal for the harmonic in the previous time frame. The difference between the actual phase and this prediction is what is transmitted for the phase of the signal for the harmonic in the particular time frame. For a particular number of binary bits to represent such harmonic, this difference prediction can be transmitted with more accuracy to the voice decoder 100 than the information representing the phase of the signal constituting such harmonic in such particular time frame. However, if the phase of the signal for such harmonic in such particular time frame 14 does not have at least the particular continuity with the phase of the signal for such harmonic in the pre-

vious time frame, the phase of the signal for such harmonic in such particular time frame is transmitted to the voice decoder 100.

As with the amplitude information, a particular number of binary bits is provided to represent the phase, or the difference prediction of the phase, for each harmonic in each time frame. The number of binary bits representing the phases, or the difference predictions of the phases, of the harmonic signals in each time frame 14 is computed as the total bits available for the time frame minus the bits already used for prior information. The phases, or the difference predictions of the phases, of the signals at the lower harmonic frequencies are indicated in a larger number of binary bits than the phases of the signals, or the difference predictions of the phases, of the signals at the higher frequencies.

The binary bits representing the phases, or the predictions of the phases, for the signals of the different harmonics in each time frame 14 are produced in a stage 130 in Figure 3, this stage being designated as "phase encoding". The binary bits representing the phases, or the prediction of the phases, of the signals at the different harmonics in each time frame 14 are transmitted through a line 132 in each time frame 14 after the binary bits representing the amplitudes of the signals at the different harmonics in each time frame.

The voice decoder 100 is shown in a simplified block form in Figure 2. The voice decoder 100 includes a line 140 which receives the coded voice signals from the voice coder 18. A transform decoder stage generally indicated at 142 operates upon these signals, which indicate the pitch frequency and the amplitudes and phases of the pitch frequency and the harmonics, to recover the signals representing the pitch frequency and the harmonics. A stage 144 performs an inverse of a Fourier frequency transform on the recovered signals representing the pitch frequency and the harmonics to restore the signals to a time domain form. These signals are further processed in the stage 144 by compensating for the effects of the Hamming window 94 shown in Figure 10. In effect, the stage 144 divides by the Hamming window 94 to compensate for the multiplication by the Hamming window in the voice coder 18. The signals in the time domain form are then separated in a stage 146 into the voice signals in the successive time frames 14 by taking account of the time overlap still remaining in the signals from the stage 144. This time overlap is indicated at 16 in Figure 6.

The transform decoder stage 142 is shown in block form in additional detail in Figure 5. The transform decoder 142 includes a stage 150 for receiving the 48, 64 or 80 bits representing the amplitudes of the pitch frequency and the harmonics and for decoding these signals to determine the amplitudes of the pitch frequency and the harmonics. In decoding such signals, the stage 150 performs a sequence of steps which are in reverse order to the steps performed during the encoding operation and which are the inverse of such steps. As a first step in such decoding, the stage 150 performs the inverse of a discrete cosine transform on such signals to obtain the frequency components of the voice signals in each time frame 14.

As will be appreciated, the number of signals produced as a result of the inverse discrete cosine transform depends upon the number of the harmonics in the voice signals at the voice coder 18 in Figure 1. The number of harmonics is then expanded or compressed to the number of harmonics at the voice coder 18 by interpolating between successive pairs of harmonics at the upper end of the frequency range. The number of harmonics in the voice signals at the voice coder 18 in each time frame can be determined in the stage 18 from the pitch frequency of the voice signals in that time frame. As will be appreciated, if an expansion in the number of harmonics occurs, the amplitude of each of these interpolated signals may be determined by averaging the amplitudes of the harmonic signals with frequencies immediately above and below the frequency of this interpolated signal.

A decompanding operation is then performed on the expanded number of harmonic signals. This decompanding operation is the inverse of the companding operation performed in the transform coder stage 26 shown in Figure 1 and in detail in Figure 3 and shown schematically in Figure 15. The decompanded signals are then restored to a base of zero (0) as a reference from the peak amplitude of all of the harmonic signals as a reference. This corresponds to a conversion of the signals from the form shown in Figure 14 to the form shown in Figure 13.

A phase decoding stage 152 (Figure 3) in Figure 5 receives the signals from the amplitude decoding stage 150. The phase decoding stage 152 determines the phases ϕ_1, ϕ_2, ϕ_3, etc. for the successive harmonics in each time frame 14. The phase decoding stage 152 does this by decoding the binary bits indicating the phase of each harmonic in each time frame 14 or by decoding the binary bits indicating the difference predictions of the phase for such harmonic in such time frame 14. When the phase decoding stage 152 decodes the difference prediction of the phase of a harmonic in a particular time frame 14, it does so by determining the phase for such harmonic in the previous time frame 14 and by modifying such phase in the particular time frame 14 in accordance with such phase prediction for such time frame.

The decoded phase signals from the phase decoding stage 152 are introduced to a harmonic reconstruction stage 154 as are the signals from the amplitude decoding stage 150. The harmonic reconstruction stage 154 operates on the amplitude signals from the amplitude decoding stage 150 and the phase signals from the phase decoding stage 154 for each time frame 14 to reconstruct the harmonic signals in such time frame. The harmonic reconstruction stage 152 reconstructs the harmonics in each time frame 152 by providing the frequency pattern (Figure 11) at different frequencies to determine the pattern at such different frequencies of the signals introduced to the stage 154.

The signals from the harmonic reconstruction stage 154 are introduced to a harmonic synthesis stage 158. The stage 158 operates to synthesize the Fourier frequency coefficients by positioning the harmonics and multiplying these harmonics by the Fourier frequency transform of the Hamming window 94 shown in Figure 10. The signals from the harmonic synthesis stage 158 pass to a stage 160 where the unvoiced signals (binary "0") in the time slots or bins 118 (Figure 16) are provided on a line 167 and are processed. In these frequency bins or slots 118, signals having a noise level represented by the average amplitude level of the harmonic signals in such time slots or bins are provided on the line 168. These signals are processed in the stage 160 to recover the frequency components in such time slots. As previously indicated, the signals from the stage 160 are subjected in the stage 144 in Figure 2 to the inverse of the Fourier frequency transform. The resultant signals are in the time domain and are modified by the inverse of the Hamming window 94 shown in Figure 10. The signals from the stage 144 accordingly represent the voice signals in the successive time frames 14. The overlap in the successive time frames 14 is removed in the stage 146 to reproduce the voice signals in a continuous pattern.

The apparatus and methods described above have certain important advantages. They employ a plurality of different techniques to determine, and then refine the determination of, the pitch frequency in each of a sequence of overlapping time frames. They employ refined techniques to determine the amplitude and phase of the pitch frequency signals and the harmonic signals in the voice signals of each time frame. They also employ refined techniques to convert the amplitude and phase of the pitch frequency signals and the harmonic signals to a binary form which accurately represents the amplitudes and phases of such signals.

The apparatus and methods described in the previous paragraph are employed at the voice coder. The voice decoder employs refined techniques which are the inverse of those, and are in reverse order to those, at the voice coder to reproduce the voice signals. The apparatus and methods employed at the voice decoder are refined in order to process, in reverse order and on an inverted basis, the encoded signals to recover the voice signals introduced to the voice encoder.

Although this invention has been disclosed and illustrated with reference to particular embodiments, the principles involved are susceptible for use in numerous other embodiments which will be apparent to persons skilled in the art. The invention is, therefore, to be limited only as indicated by the scope of the appended claims.

Claims

1. In combination for use in a voice coder to determine the pitch frequency of voice signals introduced to the voice coder,
   first means for dividing the voice signals into successive time frames,
   second means for providing a Cepstrum determination of the voice pitch frequency in the successive time frames, and
   third means for providing a harmonic gap determination of the voice pitch frequency in the successive time frames to refine the determination of the pitch frequency by the second means.

2. In a combination as set forth in claim 1,
   fourth means responsive to the detections provided by the second and third means for applying heuristic techniques to the Cepstrum determination and the harmonic gap determination for redefining the determination of the voice pitch frequency by the second and third means.

3. In a combination as set forth in claim 2 wherein
   the fourth means includes means for determining the power at low frequencies in the voice in the successive time frames and further includes means for determining the ratio of the energy at the low frequencies to the energy at the high frequencies in the successive time frames.

4. In a combination as set forth in any of claims 1-3 wherein
   the third means includes fifth means for selecting in each successive time frame a particular number of signals with the highest peak amplitudes and sixth means for determining in each successive time frame the amplitude difference between these peak amplitudes and

the troughs between these peak amplitudes and the peak amplitude of the adjacent harmonic to refine the determination of the pitch frequency by the second means.

5. In a combination as set forth in any of claims 1-4 wherein
the second means determines the locations and amplitudes of the peaks of the signals in each successive time frame.

6. In a combination as set forth in any of claims 1-5 wherein
the first through fourth means are located at a voice coding station and wherein the signals from the fourth means are transmitted to a voice decoding station and wherein
means are located at the voice decoding station to decode the transmitted signals.

7. In a combination as set forth in any of claims 1-6, wherein
means are responsive at the voice coder to the voice signals in the successive time frames for producing a frequency spectrum of the voice signals in each time frames and wherein
means are included at the voice coder for providing signals representing the amplitude of the signals in the frequency spectrum in each time frame and wherein
means are provided at the voice coder for providing signals representing the phases of the signals in the frequency spectrum in each time frame and wherein
the signals representing the pitch frequency and the amplitudes and the phases of the frequency spectrum of the voice signals in each time frame are transmitted to a voice decoding station and wherein
means are provided at the voice decoding station for operating upon the transmitted signals to recover the voice signals introduced to the voice coder.

8. In combination for use on voice signals in a voice coder,
first means for dividing the voice signals into successive time frames,
second means for converting the voice signals in each time frame into a frequency spectrum,
third means responsive to the signals from the second means for producing signals indicating the pitch frequency of the voice signals in each time frame, and
fourth means for performing additional determinations of pitch frequency on the voice signals in pairs of successive time frames, and
fifth means for interpolating the pitch frequency of the voice signals in one of the time frames in each pair in accordance with the additional determinations by the fourth means of the pitch frequency on the voice signals in the successive time frames in that pair.

9. A method as set forth in claim 8 wherein
the third means performs harmonic gap analyses and pitch match analyses on the signals from the second means in each time frame to obtain a determination of the pitch frequency of the voice signals in such time frame.

10. A method as set forth in either of claims 8 or 9 wherein
the fourth means performs a Cepstrum analysis on the voice signals in each pair of successive time frames and performs a harmonic gap analysis of the signals in each successive pair of time frames and performs an additional harmonic gap analysis of the voice signals in a particular one of each successive pair of time frames and interpolates the spectrum prior to the harmonic gap analysis in the particular one of each successive pair of time frames in accordance with the harmonic gap analysis of the signals in such successive pair of time frames.

11. A method as set forth in any of claims 8, 9 and 10, including
means for determining the amplitude and phase of each of the harmonics in each of the voice signals in each time frame,
means for converting the signals from the fourth means to a binary form for transmission, and
means for converting the determined amplitude and phase of the harmonics in each time frame to a binary form for transmission.

12. In a combination as set forth in any of claims 8-11,
a voice decoder,
means for transmitting to he voice decoder the binary signals representing the pitch frequency and the amplitude and phase of the harmonics in the voice signals in each time frame, and
means at the voice decoder for operating upon the transmitted signals to recover the voice signals introduced to the voice coder in each time frame.

13. In combination for use in a voice coder to determine the pitch frequency of voice signals in the voice coder,
first means for dividing the voice signals into successive time frames,
second means for obtaining a frequency transform of the voice signals in each of the successive time frames to obtain frequency signals in such time frame,
third means for obtaining a log spectrum of the frequency signals in each of the successive time frames,
fourth means for determining the peaks of the signals in the frequency transform in each of the successive time frames,
fifth means for determining the pitch frequencies by a harmonic gap analysis at the peaks of the peak signals in the frequency transform and at the troughs between the peak signals in each of the successive time frames, and
sixth means for determining the pitch frequency of the frequency signals in each time frame in accordance with the time between the largest peak in the voice signals in each of such time frame to refine the determination of the pitch frequency by the fifth means.

14. In a combination as set forth in claim 13 wherein
the fifth means includes means for determining the amplitudes of the signals for the frequencies around the peaks of the peak signals in each time frame and the amplitudes of the signals for the frequencies around the troughs between the peak signals in such time frame and means for averaging such determined amplitudes to select the frequency with the highest determined average amplitude.

15. In a combination as set forth in either of claims 13 or 14,
means for predicting the pitch frequency on a heuristic basis from the pitch frequencies determined by the fifth and sixth means.

16. In a combination as set forth in any of claims 13, 14 or 15 wherein
the sixth means determines the pitch frequencies by the harmonic gap analysis in the pitch frequency range of low pitch voices whether the voices are low pitch or high pitch and the sixth means also determines the pitch frequencies in the pitch frequency range of high pitch voices by the harmonic gap analysis when the voice has a high pitch.

17. In a combination as set forth in any of claims 13-16,
a voice decoder,
means for determining the amplitudes and phases of the harmonics of the pitch frequency in voice signals in each time frame,
means for transmitting to the voice decoder the signals representing the pitch frequency and the amplitudes and phases of the harmonics, and
means at the voice decoder for operating upon the signals transmitted to the voice decoder to reproduce the voice signals transmitted to the voice decoder in each time frame.

18. In combination for use in a voice coder to determine the pitch frequency of voice signals in the voice coder,
first means for dividing the voice signals into successive time frames,
second means for obtaining a frequency transform of the voice signals in each of the successive time frames to obtain a spectrum of frequency signals in such time frame,
third means for obtaining a log spectrum of the frequency signals in each time frame, and
fourth means for determining the frequency locations of the peak amplitudes and the troughs between the peak amplitudes in the spectrum of frequency signals to determine the pitch frequency in accordance with the relative differences between such peaks and troughs.

19. In a combination as set forth in claim 18 wherein
the fourth means is operative to determine the peak amplitudes of the frequency signals in each frequency transform at the frequencies of a particular number of the highest peak amplitudes in the frequency spectrum and at the frequencies around such frequencies of such peak amplitudes and the amplitudes of the signals in the frequency transforms at the frequencies of the amplitude troughs following such peak amplitudes in the frequency spectrum and at the frequencies around such troughs.

20. In a combination as set forth in either of claims 18 and 19,
fifth means for determining the pitch frequency by at least one technique other than as set forth in the fourth means, and
sixth means for selecting the pitch frequency on a heuristic basis in accordance with the determination of the pitch frequency by the

fourth and fifth means.

21. In a combination as set forth in claim 20 wherein

the sixth means includes means for determining in each time frame the pitch frequency of the voice signals in the previous time frame and for refining the determination of the pitch frequency in each time frame in accordance with the determination of the pitch frequency in the previous time frame.

22. In a combination as set forth in either of claims 20 and 21 wherein

the sixth means includes means for determining the power of the frequency signals at the low frequencies in each time frame relative to the power of the frequency signals at the high frequencies in such time frame and means for refining the determination of the pitch frequency in each time frame in accordance with such power determination.

23. In a combination as set forth in either of claims 20 or 22 wherein

the sixth means includes means for determining in each time frame the pitch frequency of the frequency signals in the previous time frame and means for determining the reliability of the determination of the pitch frequency in the previous time frame and means for refining the determination of the pitch frequency of the frequency signals in each time frame in accordance with the determination of the reliability of the determination of the pitch frequency in the previous time frame.

24. In a combination as set forth in either of claims 20 and 23,

the sixth means includes means for determining the magnitude of the low frequency power of the frequency signals in each time frame and means for determining the magnitudes of the low frequency power of the frequency signals in each time frame relative to the magnitude of the high frequency power of the frequency signals in such time frame and means for refining the determination of the pitch frequency of the frequency signals in each time frame in accordance with the determination of such magnitude of the low frequency power and the relative magnitudes of the low frequency power to the high frequency power in such time frames.

25. In a combination as set forth in any of claims 18-24 wherein

a voice decoder is included and wherein

the signals representing the pitch frequency of the voice signals in each time frame are transmitted from the voice coder to the voice decoder and wherein

signals representing the amplitudes and phases of the frequency signals in each time frame are transmitted from the voice coder to the voice decoder and wherein

means are provided at the voice decoder for operating upon the transmitted signals in each time frame to obtain the recovery of the voice signals introduced to the voice coder.

26. In combination for use on voice signals in a voice coder,

first means for dividing the voice signals into successive time frames,

second means for combining the voice signals in successive pairs of time frames to obtain an enhanced resolution of the voice signals in each time frame,

third means for obtaining a frequency transform of the voice signals into frequency signals in each of the successive pairs of time frames,

fourth means for passing the frequency signals in each of the successive pairs of time frames in a first particular range of frequencies and for providing a progressive filtering of such frequency signals for progressive frequencies above the first particular range in each of the successive pairs of time frames, and

fifth means for sub-sampling the filtered signal, and

sixth means for operating upon the signals from the fifth means for determining the pitch frequency of the frequency signals in each successive pair of time frames.

27. In a combination as set forth in claim 26,

means for providing a frequency transform of the voice signals in each time frame,

means for operating upon the signals in each time frame to produce signals representing the pitch frequency of the voice signals in that time frame, and

means for interpolating the signals representing the pitch frequency in each of the first time frames in the successive pairs of time frames with the signals representing the pitch frequency in the voice signals in such successive pair of time frames.

28. In a combination as set forth in either of claims 26 or 27,

means for determining the amplitudes of the frequency signals representing the voice signals in each successive time frame, and

means for determining the phases of the frequency signals representing the voice signals in each successive time frame.

29. In a combination as set forth in claim 28,
a voice decoder,
means for transmitting to the voice decoder the signals representing the pitch frequency and the signals representing the amplitudes and phases of the frequency signals in each time frame, and
means at the voice decoder for processing the transmitted signals to obtain a recovery of the voice signals in each time frame.

30. In a combination as set forth in any of claims 26-29,
means at the voice coder for providing a harmonic gap analysis and a harmonic difference analysis of the frequency signals in each time frame to obtain a determination of the pitch frequency in that time frame, and
means at the voice coder for providing a harmonic gap analysis and a harmonic difference analysis of the frequency signals in such successive pair of time frames to obtain a determination of the pitch frequency in such successive pair of time frames.

31. In combination for use in a voice coder to determine the pitch frequency of voice signals introduced to the voice coder,
first means for dividing the voice signals into successive time frames,
second means for providing a frequency analysis of the voice signals in each successive time frame,
third means for adding the amplitudes of the odd harmonics in the signals in the frequency transform in each time frame,
fourth means for adding the amplitudes of the even harmonics in the signals in the frequency transform in each time frame,
fifth means for normally selecting the lowest of the frequencies in the frequency transform in each time frame as the pitch frequency, and
sixth means for selecting the lowest of the frequencies in the even harmonics in each time frame as the pitch frequency when the magnitude of the addition of the amplitudes of the even harmonics in such time frame exceeds the magnitude of the addition of the amplitudes of the odd harmonies in such time frame by a particular threshold.

32. In a combination as set forth in claim 31,
means for determining the pitch frequency for the voice signals in each successive pair of time frames in accordance with a Cepstrum analysis, and
means responsive to the determination in each time frame of the pitch frequency by the sixth means and the determination of the pitch frequency in each successive pair of time frames by the Cepstrum analysis for interpolating the pitch frequency in the first one of each successive pair of time frames in accordance with the pitch frequency determined for such successive pair of time frames by the Cepstrum analysis.

33. In a combination as set forth in either of claims 31 or 32,
means for determining the pitch frequency for each time frame in accordance with a harmonic gap analysis, and
means responsive to the different determinations of the pitch frequency for each time frame for refining the selection of the pitch frequency in accordance with such different determinations.

34. In a combination as set forth in any of claims 31-33,
means for determining the reliability of the determination of the pitch frequency in the previous frame,
means for determining the cumulative magnitude in each time of the amplitudes of the signals at the low frequencies in the frequency analysis in each time frame relative to the cumulative magnitude of the amplitudes of the signals at the high frequencies in the frequency analysis in such time frame, and
means responsive to the different determinations of the pitch frequency for each time frame and to the reliability of the determination of the pitch frequency in the previous time frame and to the determination by the last mentioned means for refining the determination of the pitch frequency for each time frame.

35. In a combination as set forth in any of claims 31-34,
a voice decoder,
means for determining the amplitudes and the phases of the frequency signals in each time frame,
means for transmitting the signals representing the determined pitch frequency and the signals representing the amplitudes and phases of the frequency signals in each time frame to the voice decoder, and
means at the voice decoder for operating upon the transmitted signals for obtaining a

reproduction of the voice signals introduced to the voice decoder.

36. In combination for use on voice signals in a voice coder,

first means for dividing the voice signals into successive time frames,

second means for providing a frequency transform of the voice signals in each time frame,

third means for providing a log function of the frequency transform of the voice signals in each of the successive time frames,

fourth means for converting the log function signals in each time frame into signals having amplitudes dependent upon the amplitudes of such signals relative to the peak amplitude of the log function signal with the largest amplitude in such time frame, and

fifth means for companding the signals from the fourth means.

37. In a combination as set forth in claim 36,

sixth means for changing the number of signals from the fifth means in each time frame to a particular number, and

seventh means for obtaining a discrete cosine transform of the signals from the sixth means in each time frame.

38. In a combination as set forth in either of claims 36 or 37,

the sixth means being operative, when the number of harmonics in each time frame exceed the particular number, to eliminate every other one of the signals from the discrete cosine transform in each time frame with the highest frequencies until the particular number of signals remain.

39. In combination as set forth in any of claims 36-38,

means for converting each of the signals in the discrete cosine transform from the seventh means in each time frame into digital signals representative of the amplitudes of such signals from the seventh means wherein the number of digital signals representative of the amplitude of each of the signals from the seventh means is dependent upon the frequency of such signals.

40. In a combination as set forth in claim 39,

means for providing digital signals representing the pitch frequency of the voice signals in each time frame,

means for providing digital signals representing the phases of the frequency signals in each time frame,

a voice decoder,

means for transmitting to the voice decoder the digital signals representing the amplitudes and phases of the frequency signals in each time frame and representing the pitch frequency of such frequency signals, and

means at the voice decoder for operating upon the digital signals transmitted to the voice decoder to obtain a reproduction of the voice signals introduced to the voice coder.

41. In combination for use on voice signals in a voice coder,

first means for dividing the voice signals into successive time frames,

second means for converting the voice signals in each time frame into signals representing the different frequencies in such time frame,

third means for emphasizing in each time frame the frequency signals with peak amplitudes relative to the frequency signals with low amplitudes,

fourth means for limiting the number of signals at the high frequencies in each time frame to reduce the number of the frequency signals in such time frame to a particular value,

fifth means for producing signals representing a frequency transform of the frequency signals from the fourth means, and

sixth means for converting the transformed signals from the fifth means to digital signals representative of the amplitude of such signals.

42. In a combination as set forth in claim 41,

the fifth means providing a discrete cosine transform of the signals from the fourth means, and

the sixth means providing a greater number of digital signals to represent the amplitudes of the signals at low frequencies from the fifth means than the number of signals to represent the amplitudes of the signals at high frequencies from the fifth means.

43. In a combination as set forth in either of claims 41 or 42,

means for determining the pitch frequency of the voice signals from the first means in each time frame, and

means for determining the phases of the frequency signals in each time frame.

44. In a combination as set forth in any of claims 41-43,

means for determining the pitch frequency of the voice signals in each time frame by at

least two (2) independent analyses and for using heuristic techniques on the pitch frequencies determined in such time frame by at least such two (2) independent analyses to provide a refined determination of the pitch frequency in such time frame, and

means for converting the refined determination of the pitch frequency for the voice signals in each time frame into digital signals representing the pitch frequency, and

means for determining the phases of the frequency component signals in each time frame.

45. In a combination as set forth in claims 41-44 wherein

a voice decoder is provided and wherein

the digital signals representative of the amplitudes and phases of the frequency signals in each time frame and representing the pitch frequency in such time frame are transmitted to the voice decoder and wherein

means are provided at the voice decoder for operating upon the digital signals in each time frame to obtain a reproduction of the voice signals introduced to the voice coder.

46. In a combination as set forth in any of claims 41-45 wherein

the fifth means include means for determining the pitch frequency of the voice signals in each time frame by at least two of a Cepstrum analysis, a harmonic gap analysis, a pitch match analysis and a harmonic difference analysis.

47. In combination for use on voice signals in a voice encoder,

first means for separating the voice signals into successive time frames,

second means for transforming the voice signals in each successive time frame into frequency signals representative of the voice signals in such time frame,

third means for determining the pitch frequency of the frequency signals in each time frame and for producing digital signals representing such pitch frequency,

fourth means for determining the amplitudes of the frequency signals in each time frame and for producing digital signals representing such amplitudes,

fifth means for determining the phases of the frequency signals in each time frame and for producing signals representing such phases,

sixth means for determining the continuity of the pitch frequency in the successive time frames,

seventh means for providing a prediction of the difference in the phases of the frequency signals in the successive time frames when the pitch frequencies of the voice signals in such time frame and the adjacent time frames have continuities within particular limits and for producing signals presenting such predictions, and

eighth means for converting the signals representing the phases, and the predictions of the phases, of the frequency signals in each time frame into digital signals representing such phases and such predictions.

48. In a combination as set forth in claim 47,

the third means including:

ninth means for providing more than one different type of determination of the pitch frequency of the voice signals in each time frame, and

tenth means for using heuristic techniques on the different determinations of the pitch frequency in each time frame to provide a refined determination of the pitch frequency in each time frame.

49. In a combination as set forth in either of claims 47 or 48,

the fifth means being operative in each time frame to emphasize the frequency signals with high amplitudes relative to the frequency signals with low amplitudes during the determination of the amplitudes of the frequency signals in such time frame and to produce signals representing such emphasized amplitudes and being farther operative to emphasize the amplitudes of the frequency signals with low frequencies relative to the amplitudes of the frequency signals with high frequencies.

50. In a combination as set forth in any of claims 47-49,

a voice decoder,

means for transmitting the digital signals from the third, fourth, fifth and eighth means in each time frame to the voice decoder, and

means at the voice decoder for operating upon the transmitted signals to recover the voice signals in the successive time frames.

51. In combination for use in voice signals in a voice encoder,

first means for separating the voice signals into successive time frames,

second means for transforming the voice signals in each successive time frame into frequency signals representative of the voice

signals in such time frame,

third means for providing a refined determination of the pitch frequency of the frequency signals in each time frame by at least two (2) different analyses and for producing digital signals representing such pitch frequency,

fourth means for determining the disposition of each harmonic in the frequency signals in individual ones of a plurality of time blocks and in individual ones of a plurality of grids within each time block,

fifth means for determining the phases of the frequency signals in each time frame in accordance with the determination by the fourth means and for producing digital signals representing such phases,

sixth means for determining the amplitudes of the frequency signals in each time frame in accordance with the determinations by the fourth means and for producing digital signals representing such determinations, and

seventh means for transmitting the digital signals from the third, fifth and sixth means.

52. In a combination as set forth in claim 51 wherein

the third means provides the refined determination of the pitch frequency by providing at least two (2) of a harmonic gap analysis, a Cepstrum analysis, a pitch match analysis and a harmonic difference analysis.

53. In a combination as set forth in either of claims 51 and 52,

the sixth means including means for limiting the number of frequency signals to a particular value and including means for taking a discrete cosine transform of the frequency signals.

54. In a combination as set forth in any of claims 51-53,

a voice decoder, and

means at the voice decoder for processing the transmitted digital signals to recover the voice signals in the successive time frames.

55. In combination for use in a voice decoder to recover voice signals introduced to the voice coder where the voice signals are processed in successive time frames and wherein the voice signals in each time frame are subjected to a first frequency transform to produce frequency signals in each time frame and where inversion signals are produced representing the difference between the peak amplitude of the frequency signals in each time frame and the amplitude of the frequency signals in such time frame and where the amplitudes of the inversion signals are companded and wherein a second frequency transform is performed on the companded signals and wherein the amplitudes of the signals in the second frequency transform are converted to digital signals,

first means for decoding the digital signals representing the signals in the second frequency transform in each time frame,

second means for providing an inverse frequency transform of the signals from the first means in each time frame,

third means for decompanding the signals from the second means in each time frame, and

fourth means for inverting the decompanded signals in each time frame relative to the peak amplitudes of the voice signals in such time frame.

56. In a combination as set forth in claim 55 wherein

after the companding operation, the number of frequency harmonics in each time frame is limited or expanded at the voice coder to a particular value. by eliminating or adding particular ones of the frequency signals at the high frequencies,

the third means being operative to decompand the limited number of frequency signals.

57. In a combination as set forth in either of claims 55 or 56 wherein

the voice coder provides voice signals in particular time blocks in each time frame and unvoiced signals in the other time blocks in each time frame and wherein means are provided at the voice decoder for synthesizing the signals from the decoder to determine the amplitudes of the harmonic signals in the voiced and unvoiced time blocks in each time frame.

58. In a combination as set forth in any of claims 55-57 wherein

signals are provided at the voice coder to represent the phases of the frequency signals in each time frame and wherein means are provided at the voice decoder for restoring the voice signals in each time frame in accordance with the pitch frequency and the signals representing the amplitudes and phases of the frequency signals in each time frame.

59. In a combination as set forth in any of claims 58 wherein

the time frames at the voice coder are overlapped and wherein a frequency transform

is provided on the voice signals in each time at the voice coder and wherein an inverse Fourier frequency transform is performed at the voice decoder on the signals representing the restored frequency signals in each time frame to recover the voice signals in the successive time frames and wherein the overlap in the recovered voice signals in the successive time frames is removed to recover the voice signals.

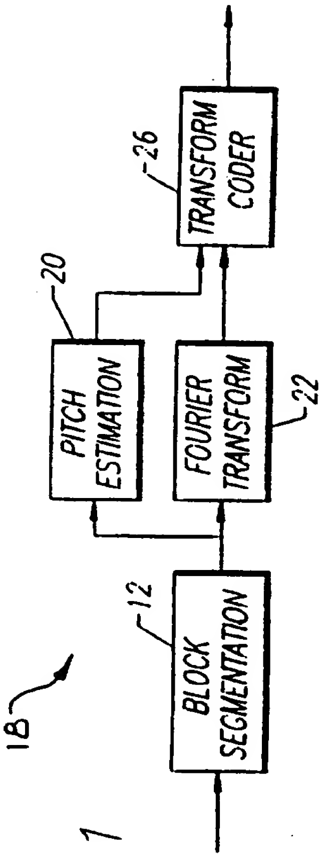


FIG. 1

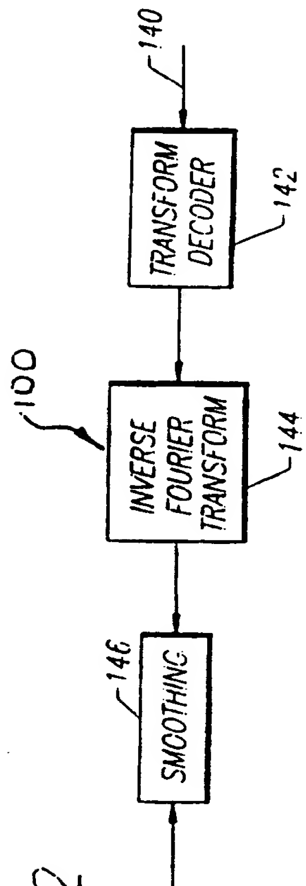


FIG. 2

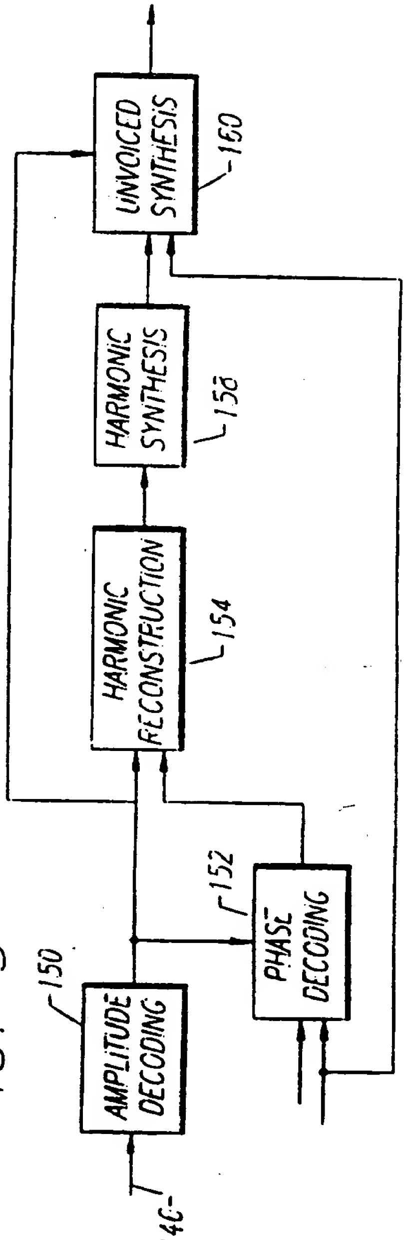


FIG. 5

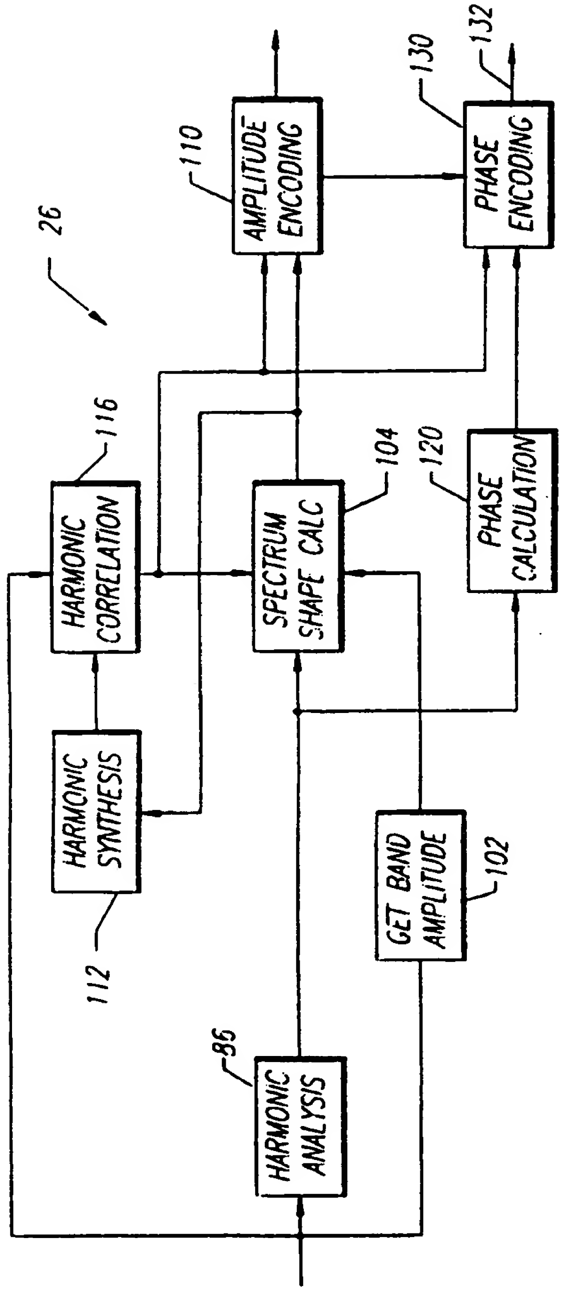

26
HARMONIC ANALYSIS
86
HARMONIC SYNTHESIS
112
HARMONIC CORRELATION
116
GET BAND AMPLITUDE
102
SPECTRUM SHAPE CALC
104
PHASE CALCULATION
120
AMPLITUDE ENCODING
110
PHASE ENCODING
130
132

FIG. 3

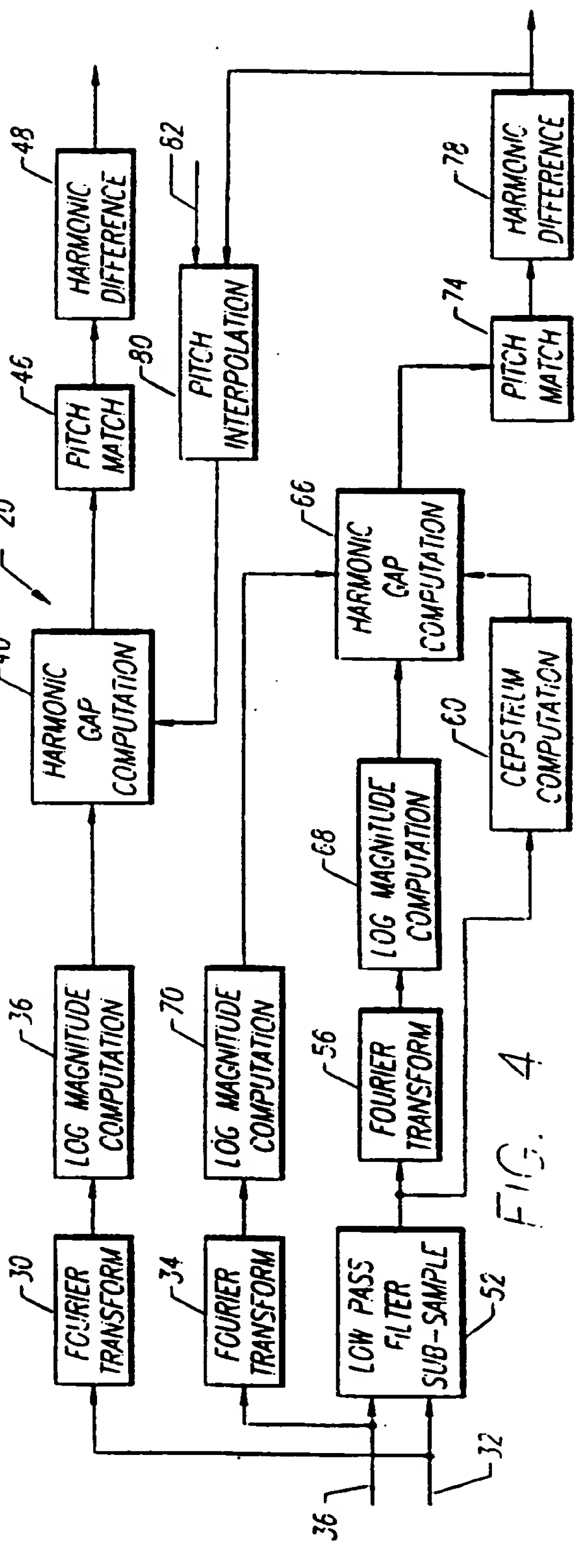

20
36
32
FOURIER TRANSFORM
30
FOURIER TRANSFORM
34
LOW PASS FILTER SUB-SAMPLE
52
LOG MAGNITUDE COMPUTATION
36
LOG MAGNITUDE COMPUTATION
70
FOURIER TRANSFORM
56
LOG MAGNITUDE COMPUTATION
68
CEPSTRUM COMPUTATION
60
HARMONIC GAP COMPUTATION
40
HARMONIC GAP COMPUTATION
66
PITCH MATCH
46
PITCH INTERPOLATION
80
PITCH MATCH
74
HARMONIC DIFFERENCE
48
62
HARMONIC DIFFERENCE
78

FIG. 4

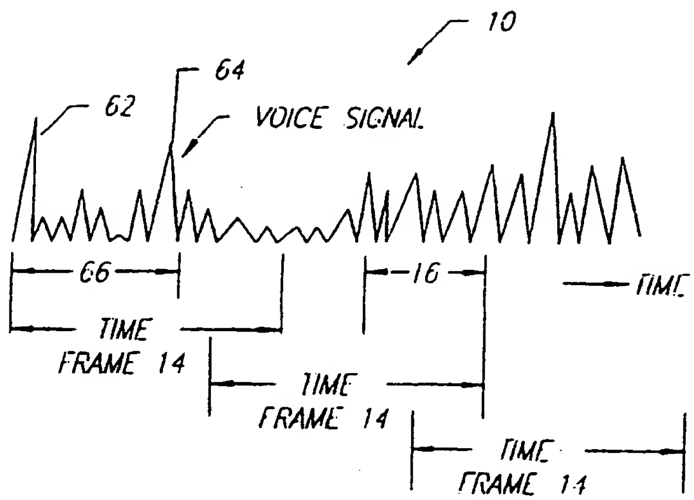
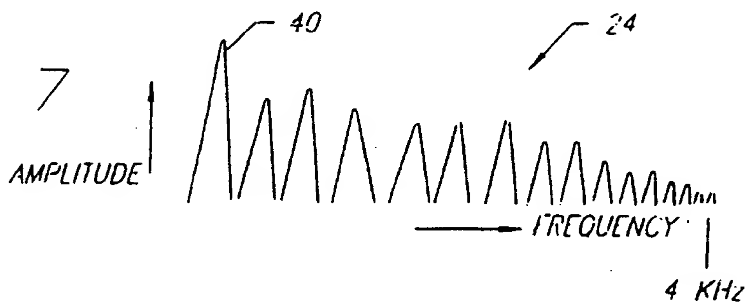


FIG. 6

FIG. 7



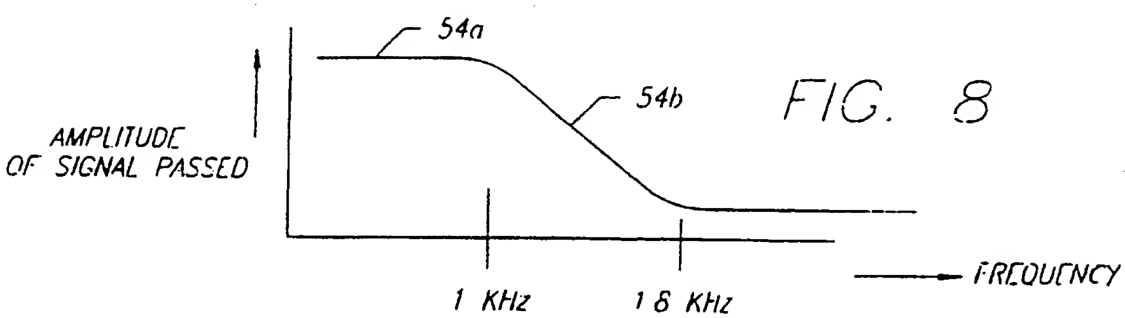
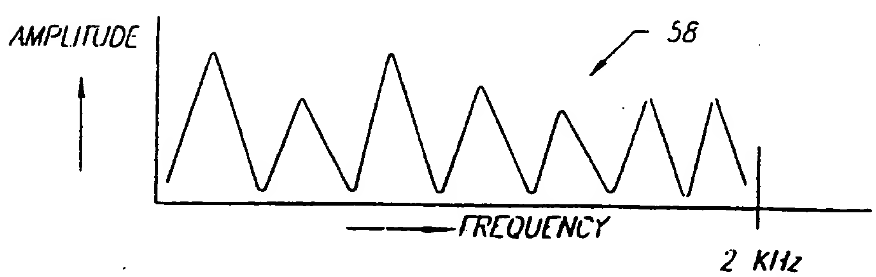


FIG. 8

FIG. 9



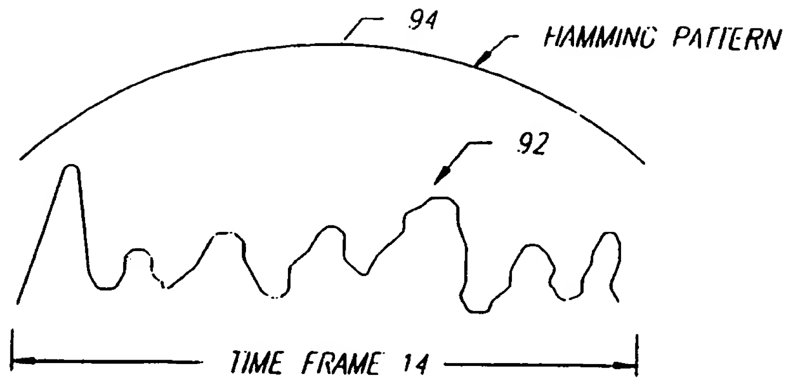


FIG. 10

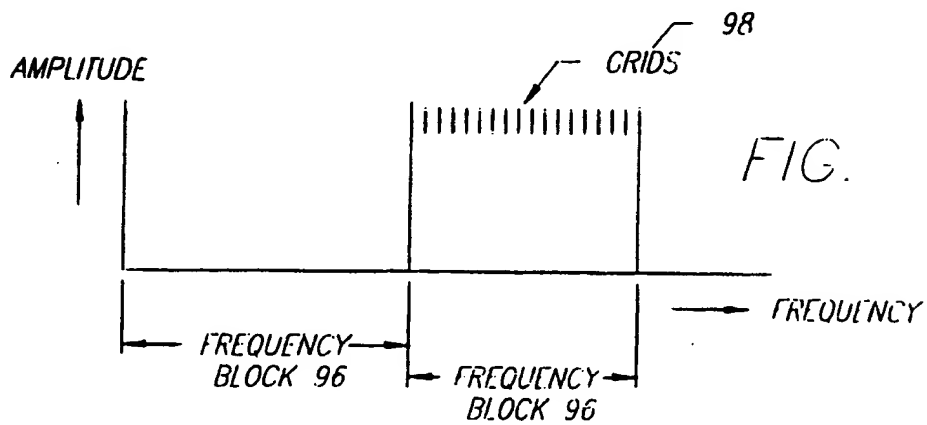


FIG. 12

FIG. 11

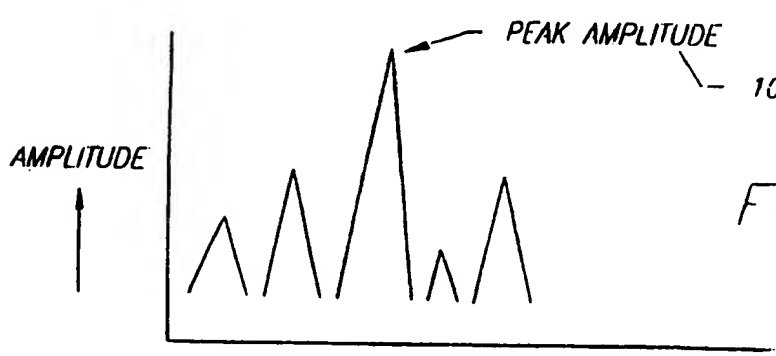


FIG. 13

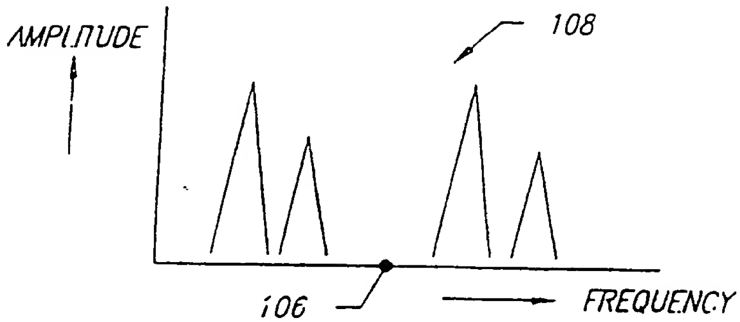


FIG. 14

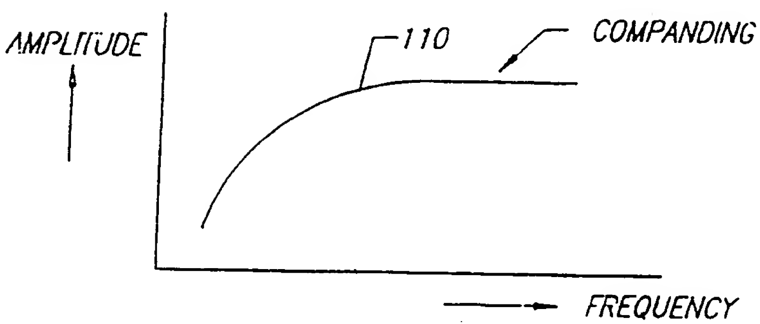


FIG. 15

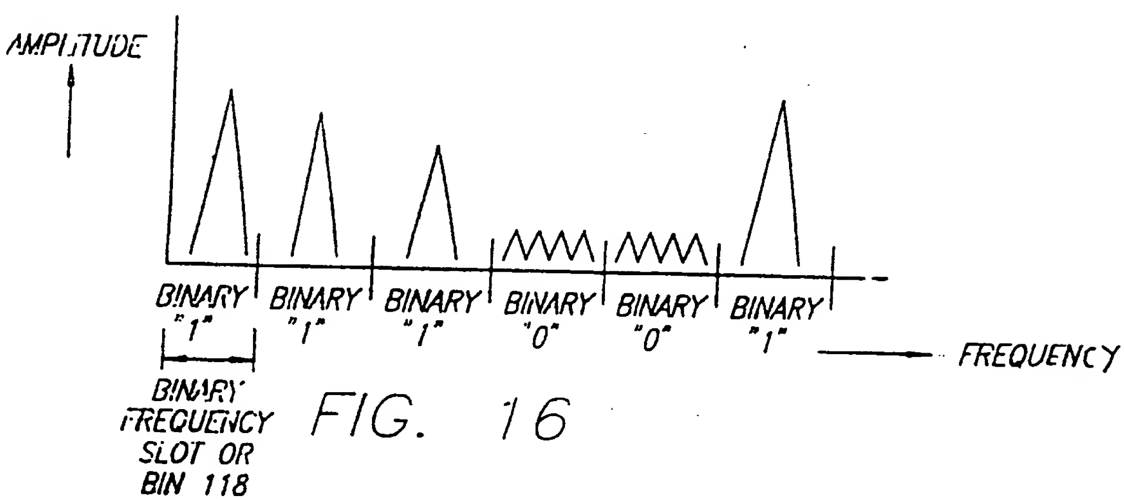


FIG. 16

(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) Publication number: **0 538 877 A3**

(12) EUROPEAN PATENT APPLICATION

(21) Application number: **92118176.4**

(51) Int. Cl.[5]: **G10L 7/06**

(22) Date of filing: **23.10.92**

(30) Priority: **25.10.91 US 782669**

(43) Date of publication of application: **28.04.93 Bulletin 93/17**

(84) Designated Contracting States: **CH DE FR GB IT LI SE**

(88) Date of deferred publication of the search report: **09.02.94 Bulletin 94/06**

(71) Applicant: **MICOM COMMUNICATIONS CORP.**
4100 Los Angeles Avenue
Simi Valley, California 93063(US)

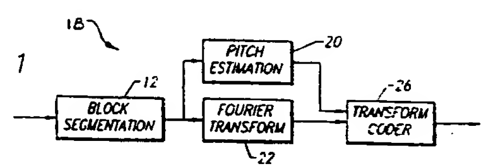
(72) Inventor: **Jain,Jaswant R.**
21638 Wohelo Trail
Chatsworth, California 91311(US)

(74) Representative: **Patentanwälte Grünecker, Kinkeldey, Stockmair & Partner**
Maximilianstrasse 58
D-80538 München (DE)

(54) Voice coder/decoder and methods of coding/decoding.

(57) A new adaptive Fourier transform coder/decoder encodes periodic components of speech signals and decodes the encoded periodic components. The pitch frequency of voice signals in successive time frames at the voice coder may be determined as by (1) Cepstrum analysis (e.g. the time between successive peak amplitudes in each time frame), (2) harmonic gap analysis (e.g. the amplitude differences between the peaks and troughs of the peak amplitude signals of the frequency spectrum) (3) harmonic matching, (4) filtering of the frequency signals in successive pairs of time frames and the performance of (1), (2) and (3) on the filtered signals to provide pitch interpolation on the first frame in the pair and (5) pitch matching. The amplitude and phase of the pitch frequency and harmonic signals are determined by techniques refined relative to the prior art to provide amplitude and phase signals with enhanced resolution. Such amplitudes may be converted to a simplified digital form by (a) taking the logarithm of the frequency signals, (b) selecting the signal with the peak amplitude, (c) offsetting the amplitudes of the logarithmic signals relative to such peak amplitude, (d) companding the offset signals, (e) reducing the number of harmonics to a particular limit by eliminating alternate high frequency harmonics, (f) taking a discrete cosine transform of the remaining signals and (g) digitizing the transformed signals. If the pitch frequency has a continuity within particular limits in successive time frames, the phase difference of the signals between successive time frames is provided. At a displaced voice decoder, the signal amplitudes are determined by performing, in order, the inverse of steps (g) through (a). These signals and the signals representing pitch frequency and phase are processed to recover the voice signals.

FIG. 1



EP 0 538 877 A3

European Patent Office

EUROPEAN SEARCH REPORT

Application Number

EP 92 11 8176

DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (Int. Cl.5)
A	ICASSP'90 (1990 INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING, Albuquerque, New Mexico, 3rd - 6th April 1990), vol. 1, pages 253-256, IEEE, New York, US; M. SCOTT ANDREWS et al.: "Robust pitch determination via SVD based cepstral methods" * Paragraph III: "An integrated music-cepstrum algorithm" * ---	1,5	G 10 L 7/06
A	THE JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA, vol. 65, no. 1, January 1979, pages 223-228, New York, US; T.V. SREENIVAS et al.: "Pitch extraction from corrupted harmonics of the power spectrum" * Paragraph I: "Logical filtering" * ---	1,4,9, 13-16, 18,19	
A	ICASSP'90 (1990 INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING, Albuquerque, New Mexico, 3rd - 6th April 1990) vol. 1, pages 17-20, IEEE, New York, US; J.S. MARQUES et al.: "Harmonic coding at 4.8 Kb/s" * Paragraph 3: "Fundamental frequency estimation"; paragraph 4: "Harmonic coding"; paragraph 5: "Phase quantization" * ---	1,4,6,7 ,9,11-19,25, 28,29, 35,40, 43,45,	TECHNICAL FIELDS SEARCHED (Int. Cl.5) G 10 L 7/06

-/-

The present search report has been drawn up for all claims

Place of search	Date of completion of the search	Examiner
THE HAGUE	01-12-1993	ARMSPACH J F A

CATEGORY OF CITED DOCUMENTS

X : particularly relevant if taken alone
Y : particularly relevant if combined with another document of the same category
A : technological background
O : non-written disclosure
P : intermediate document

T : theory or principle underlying the invention
E : earlier patent document, but published on, or after the filing date
D : document cited in the application
L : document cited for other reasons

& : member of the same patent family, corresponding document

European Patent Office

EUROPEAN SEARCH REPORT

Application Number

EP 92 11 8176

DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (Int. Cl.5)
A	ICASSP'88 (1988 INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH AND SIGNAL PROCESSING, New York, 11th - 14th April 1988), vol. 1, pages 537-540, IEEE, New York, US; K. MIN et al.: "Automated two speaker separation system" * Pages 538-539: "Pitch detection"; figure 3 * ---	8,21,23 ,27	
A	EP-A-0 260 053 (AMERICAN TELEPHONE AND TELEGRAPH CO.) * Figure 1 * ---	8,9,18, 20	
A	IEEE TRANSACTIONS ON ACOUSTICS, SPEECH AND SIGNAL PROCESSING, vol. ASSP-29, no. 4, August 1981, pages 786-794, New York, US; D.B. PAUL: "The spectral envelope estimation vocoder" * Paragraph III: "The spectral envelope estimation algorithm" * ---	13,18, 51	TECHNICAL FIELDS SEARCHED (Int. Cl.5)
A	THE JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA, vol. 83, no. 1, January 1988, pages 257-264, New York, US; D.J. HERMES: "Measurement of pitch by subharmonic summation" * Figure 1 * --- -/-	31	

The present search report has been drawn up for all claims

Place of search	Date of completion of the search	Examiner
THE HAGUE	01-12-1993	ARMSPACH J F A

CATEGORY OF CITED DOCUMENTS

X : particularly relevant if taken alone
Y : particularly relevant if combined with another document of the same category
A : technological background
O : non-written disclosure
P : intermediate document

T : theory or principle underlying the invention
E : earlier patent document, but published on, or after the filing date
D : document cited in the application
L : document cited for other reasons

& : member of the same patent family, corresponding document

European Patent Office

EUROPEAN SEARCH REPORT

Application Number

EP 92 11 8176

DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (Int. Cl.5)
A	ICASSP'83 (IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH AND SIGNAL PROCESSING, Boston, Massachusetts, 14th - 16th April 1983), vol. 2, pages 471-474, IEEE, New York, US; T.A. RICE et al.: "Parallel processing for computationally intensive speech analysis operations" * Page 471, right-hand column, lines 22-27 * ---	32,47	
A	EP-A-0 259 950 (AMERICAN TELEPHONE AND TELEGRAPH CO.) * Abstract * ---	37,40	
A	IEEE TRANSACTIONS ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING, vol. 39, no. 2, February 1991, pages 538-541, New York, US; F. WANG et al.: "Cepstrum analysis using discrete trigonometric transforms" * Abstract * ---	42	TECHNICAL FIELDS SEARCHED (Int. Cl.5)
A	ICASSP'85 (IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING, Tampa, Florida, 26th - 29th March 1985), vol. 3, pages 945-948, IEEE, New York, US; R.J. McAULAY et al.: "Mid-rate coding based on a sinusoidal representation of speech" * Page 947, left-hand column, lines 43-48 * ---	37,40	
A	EP-A-0 337 636 (AMERICAN TELEPHONE AND TELEGRAPH CO.) * Page 5, lines 54-58 * --- -/-	47	

The present search report has been drawn up for all claims

Place of search	Date of completion of the search	Examiner
THE HAGUE	01-12-1993	ARMSPACH J F A

CATEGORY OF CITED DOCUMENTS

X : particularly relevant if taken alone
Y : particularly relevant if combined with another document of the same category
A : technological background
O : non-written disclosure
P : intermediate document

T : theory or principle underlying the invention
E : earlier patent document, but published on, or after the filing date
D : document cited in the application
L : document cited for other reasons

& : member of the same patent family, corresponding document

European Patent Office

EUROPEAN SEARCH REPORT

Application Number

EP 92 11 8176

DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (Int. Cl.5)
A	IEEE TRANSACTIONS ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING, vol. 36, no. 8, August 1988, pages 1223-1235, New York, US; D.W. GRIFFIN et al.: "Multiband excitation vocoder" * Figure 10 *	47	
A	US-A-4 829 574 (THE UNIVERSITY OF MELBOURNE) * Claim 1 *	41	
A	EUROSPEECH'89 (EUROPEAN CONFERENCE ON SPEECH COMMUNICATION AND TECHNOLOGY, Paris, 26th - 29th September 1989), vol. 1, pages 466-469, CEP consultants, Edinburgh, GB; T.J. MOULSLEY et al.: "An adaptive voiced/unvoiced speech classifier" * Page 468, left-hand column, lines 1-7 *	3	

TECHNICAL FIELDS SEARCHED (Int. Cl.5)

The present search report has been drawn up for all claims

Place of search	Date of completion of the search	Examiner
THE HAGUE	01-12-1993	ARMSPACH J F A

CATEGORY OF CITED DOCUMENTS

X : particularly relevant if taken alone
Y : particularly relevant if combined with another document of the same category
A : technological background
O : non-written disclosure
P : intermediate document

T : theory or principle underlying the invention
E : earlier patent document, but published on, or after the filing date
D : document cited in the application
L : document cited for other reasons

& : member of the same patent family, corresponding document